WAVEFM: WAVELET DECOMPOSITION MASKED RECONSTRUCTION FOR MULTI-SCALE TIME-SERIES FOUNDATION MODEL

Seongbeom Park[1]*, Hyunwoo Seo[1]*, Yichen Ma[2]*, Hojin Cho[1], Chiehyeon Lim[1,3], Jianjun Shi[2]
[1] Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea
[2] Georgia Institute of Technology, Atlanta, Georgia, USA
[3] POSCO N.EX.T Hub, POSCO Holdings Inc., Seoul, Republic of Korea
{sbpark1021, hseo75, chjin314, chlim}@unist.ac.kr
{yma447, jianjun.shi}@gatech.edu

ABSTRACT

Encoder-only Time-Series Foundation Models (TSFMs) learn to understand time series through time-domain masked reconstruction. This objective captures transferable temporal dynamics, making them a strong universal backbone across domains and tasks. However, simple time-domain masked reconstruction does not fully capture the underlying temporal dynamics. In particular, it does not explicitly reflect the multi-scale nature of real-world time series and can bias learning toward low-frequency trends, reducing sensitivity to fine-grained temporal changes. To address this, we propose WaveFM, which decomposes each time series into scale-separated components and is trained to model within-scale temporal structure and cross-scale composition back to the original time series signal. Therefore, WaveFM explicitly emphasizes key features for understanding time series, learning how coarse trends and fine-scale deviations jointly explain the time series. Experiments on long-horizon forecasting and imputation show consistent improvements over encoder-only TSFM baselines, supporting the importance of explicitly decomposing multi-scale components for understanding time series.

Track: Research

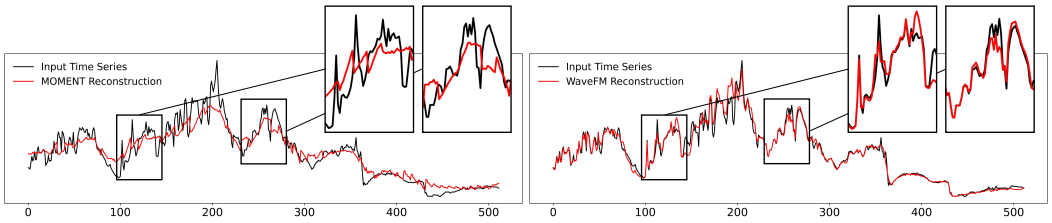

Figure 1: Reconstruction results comparing our method with MOMENT (Goswami et al., 2024) on the Weather dataset. WaveFM more faithfully reconstructs both coarse low-frequency structure and fine-grained fluctuations, indicating a stronger understanding of underlying temporal dynamics.

1 INTRODUCTION

Encoder-only time-series foundation models (TSFMs) learn transferable representations by reconstructing masked segments from surrounding context, enabling the capture of intrinsic temporal dependencies and dynamics (Woo et al., 2024; Prabhakar Kamarthi & Prakash, 2024). Through large-scale pre-training under this masked modeling paradigm, these models acquire general-purpose representations that adapt efficiently to diverse downstream tasks with minimal task-specific fine-tuning. Compared to encoder–decoder or decoder-only backbones (Wang et al., 2025a; Zhang et al.,

*Equal contribution.

2025), encoder-only TSFMs offer a compact and generalizable pre-training framework, making them a promising foundation for universal time series analysis.

Despite their empirical success, most encoder-only TSFMs rely on time-domain masked reconstruction as the sole pre-training objective (Figure 1), recovering missing values directly in the original signal space. However, real-world time series are inherently multi-scale: coarse low-frequency components coexist with fine high-frequency fluctuations and residual dynamics. Because low-frequency components typically dominate signal energy, time-domain reconstruction tends to bias learning toward smoother, large-scale patterns while under-emphasizing subtle yet informative high-frequency variations (Han et al., 2025). More fundamentally, multi-scale components are not independent (Li et al., 2025); they form a residual hierarchy whose interactions collectively compose the observed signal. However, time-domain masking provides no explicit supervision for disentangling scale-specific information or modeling cross-scale interactions, leaving compositional structure largely implicit and limiting representation expressiveness.

To address these limitations, we propose WaveFM, which learns to understand time series through (i) wavelet-decomposed multi-scale components and (ii) their composition into the original signal. Decomposition has been shown to reveal temporal patterns by separating signals across characteristic scales (Wu et al., 2021). In particular, the wavelet transform provides a principled multi-scale representation, expressing a time series as an approximation coefficient and a hierarchy of residual detail coefficients (Mallat, 1989). Its invertible structure yields a lossless mapping between coefficients and the original time series, making it well-suited for learning scale-wise information and cross-scale composition.

Building on this property, we propose a **scale coefficient attention (SCA)** mechanism that captures (i) hierarchical dependencies across scales and (ii) temporal dynamics within each scale. **Masked coefficient reconstruction** trains the model to recover missing coefficients at each scale, providing supervision for scale-wise disentanglement and residual composition. This approach not only complements the time-domain reconstruction but also achieves a comprehensive understanding of time series by modeling the composition among multi-scale components. Lastly, **scale energy ratio distribution calibration** guides the model to capture the scale-wise energy distribution of the time series. Together, these components introduce explicit self-supervisory signals for multi-scale disentanglement and cross-scale residual composition, mitigating over-smoothing of high-frequency dynamics and yielding more transferable representations for downstream tasks. Our experiments on long-horizon forecasting and imputation show consistent improvements over encoder-only TSFM baselines, highlighting the importance of explicit multi-scale decomposition.

2 WaveFM Methodology

2.1 Overview

Figure 2 summarizes WaveFM, which pre-trains an encoder-only TSFM by masked reconstruction in the wavelet domain. Given an input window, a multi-level wavelet transform is applied to obtain one low-frequency approximation stream and multiple high-frequency detail streams. Each coefficient stream is patchified, embedded, and randomly masked. The resulting tokens are processed by scale-coefficient attention blocks that model interactions across scales and temporal dependencies within each scale. The model reconstructs the masked coefficient patches and maps them back to the time domain via the inverse wavelet transform. WaveFM is pre-trained under the three objectives: masked coefficient reconstruction, time-domain reconstruction, and scale energy ratio distribution calibration that learns the wavelet energy allocation of the input window.

2.2 Wavelet Transform and Tokenization

For each input window $\mathbf{x} \in \mathbb{R}^T$, apply an L-level wavelet transform $\mathcal{W}_L$ and obtain an approximation coefficient vector $\mathbf{a}^{(L)}$ and detail coefficient vectors $\{\mathbf{d}^{(\ell)}\}_{\ell=1}^{L}$. The set of coefficient streams is denoted as $\mathcal{S} = \{a^{(L)}, d^{(L)}, \ldots, d^{(1)}\}$ and write $\{\mathbf{c}_s\}_{s\in\mathcal{S}} = \mathcal{W}_L(\mathbf{x})$. The inverse wavelet transform satisfies $\mathbf{x} = \mathcal{W}_L^{-1}(\{\mathbf{c}_s\}_{s\in\mathcal{S}})$. Each stream $s \in \mathcal{S}$ is patchified into N patches using a scale-dependent patch length $P_s = |\mathbf{c}_s|/N$, so that the number of patches is consistent across scales. Let $\mathbf{p}_{s,i} = \mathcal{P}_s(\mathbf{c}_s)_i$ for $i = 1, \ldots, N$ denote the i-th patch from stream s. Each

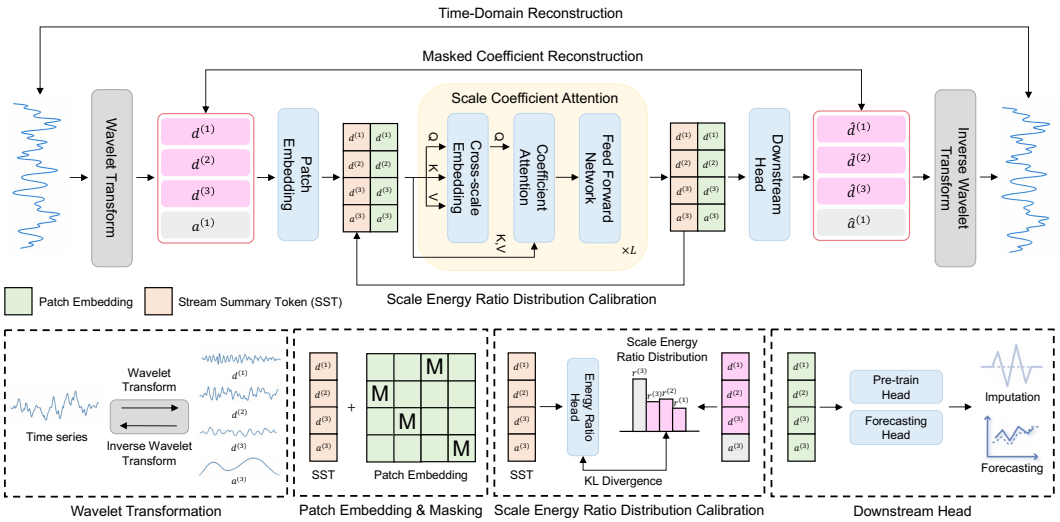

Figure 2: Overview of the proposed WaveFM pre-training framework.

patch is embedded with a stream-specific embedding layer E_s as $\mathbf{z}_{s,i} = E_s(\mathbf{p}_{s,i}) \in \mathbb{R}^D$. Random masking is applied independently per stream by setting the embeddings of masked patches to zero. Also, a learnable stream summary token (SST) is added to every stream to learn how much each stream contributes to compose the time series. The SST token acts as a stream-level summary representation, denoted as $\mathbf{z}_{s,0} = \mathbf{z}_{\text{SST}}^{(s)} \in \mathbb{R}^D$. The final token sequence for stream s is $\tilde{\mathbf{Z}}_s = [\tilde{\mathbf{z}}_{s,0}; \tilde{\mathbf{z}}_{s,1}; \ldots; \tilde{\mathbf{z}}_{s,N}] \in \mathbb{R}^{(N+1)\times D}$. Additional background on wavelet transform is provided in Appendix A.

2.3 Scale Coefficient Attention

To understand time series, a model should capture two complementary structures within time series: (1) compositional dependencies across multi-scale patterns and (2) temporal dependencies within each scale. Since each coefficient stream provides only a partial context of the underlying signal, patches across coefficients are statistically dependent; their dependencies must be integrated to reconstruct the time series. The model is therefore required to learn how coefficients jointly compose the time series, while also modeling temporal dynamics within each scale. Thus, we propose a scale coefficient attention (SCA) mechanism, consisting of (i) cross-scale embedding and (ii) coefficient attention. A sequence of patches across scales at the patch index i is defined as $\tilde{\mathbf{Z}}_i = \left[\tilde{\mathbf{z}}_{s,i}\right]_{s\in\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|\times D}, i = 0, \ldots, N$. Then, the SCA mechanism is formulated as

$$\mathbf{H}_i^{\text{scale}} = \text{CrossScaleEmb}\left(\tilde{\mathbf{Z}}_i\right) = \text{MHA}\left(\tilde{\mathbf{Z}}_i, \tilde{\mathbf{Z}}_i, \tilde{\mathbf{Z}}_i\right) \in \mathbb{R}^{|\mathcal{S}|\times D}. \tag{1}$$

$$\mathbf{H}_s^{\text{coef}} = \text{MHA}\left(\mathbf{H}_s^{\text{scale}}, \tilde{\mathbf{Z}}_s, \tilde{\mathbf{Z}}_s\right) \in \mathbb{R}^{(N+1)\times D}. \tag{2}$$

where $\mathbf{H}_s^{\text{scale}} \in \mathbb{R}^{(N+1)\times D}, \forall s \in \mathcal{S}$. Specifically, SCA constructs a cross-scale embedding that aggregates information across resolutions and encodes inter-scale relations by applying self-attention to $\tilde{\mathbf{Z}}_i$, yielding $\mathbf{H}_i^{\text{scale}}$. The outputs $\mathbf{H}_i^{\text{scale}}$ are rearranged into per-stream sequences $\mathbf{H}_s^{\text{scale}}$, providing each token with inter-scale contexts. Then, the coefficient attention integrates the multi-scale and within-scale dependencies encoded in wavelet coefficients. In particular, the coefficient attention defines the per-stream sequence of the cross-scale embeddings as queries and the original stream tokens $\tilde{\mathbf{Z}}_s$ as keys and values. It produces contextualized representations that align cross-scale contexts in $\mathbf{H}_s^{\text{scale}}$ with stream-specific content in $\tilde{\mathbf{Z}}_s$, thereby capturing the compositional structure of the underlying temporal dynamics.

Table 1: **Long-horizon forecasting.** MAE/MSE averaged over horizons {96,192,336,720}.

Dataset	WaveFM-small		WaveFM-base		MOMENT		LPTM	
	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
ETTh1	0.4424	0.4275	**0.4374**	**0.4189**	0.4529	0.4442	0.4690	0.4740
ETTh2	0.3919	0.3492	**0.3909**	**0.3446**	0.4006	0.3585	0.4029	0.3675
ETTm1	0.3886	0.3733	0.3876	0.3701	0.3952	0.3798	**0.3869**	**0.3653**
ETTm2	0.3177	0.2586	**0.3153**	**0.2565**	0.3246	0.2627	0.3216	0.2622
Electricity	0.2594	0.1653	**0.2590**	**0.1653**	0.2987	0.1973	0.2724	0.1739
Weather	0.2778	0.2401	0.2789	0.2411	0.2871	0.2472	**0.2776**	**0.2375**

Table 2: **Imputation.** MAE/MSE averaged over mask ratios $\{0.125, 0.25, 0.375, 0.5\}$.

Dataset	WaveFM-small		WaveFM-base		MOMENT	
	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	0.234	0.223	**0.206**	**0.217**	0.698	0.581
ETTh2	0.074	0.132	**0.065**	**0.126**	0.287	0.370
ETTm1	**0.173**	0.192	0.189	**0.190**	0.405	0.423
ETTm2	**0.058**	**0.112**	0.059	**0.112**	0.143	0.255
Electricity	0.179	0.246	**0.156**	**0.224**	0.181	0.307
Weather	**0.052**	**0.091**	0.056	0.092	0.096	0.157

2.4 Scale Energy Ratio Distribution

Time series are composed of multi-scale components, and the relative importance of each coefficient stream to the original signal can vary substantially across windows. Without an explicit constraint, masked reconstruction can over-rely on dominant low-frequency streams and fail to leverage fine-scale details. To capture and calibrate this scale-allocation behavior, the wavelet energy ratio distribution is used as a compact, window-level summary of how signal energy is allocated across coefficient streams. For each stream $s \in \mathcal{S}$ with coefficients $\mathbf{c}_s$, the energy ratio is defined as $r_s = \|\mathbf{c}_s\|_2^2 \big/ \sum_{s\in\mathcal{S}} \|\mathbf{c}_s\|_2^2$, and collect $\mathbf{r} = (r_s)_{s\in\mathcal{S}}$ with $\sum_{s\in\mathcal{S}} r_s = 1$. By Parseval's theorem for orthonormal wavelets, r_s is the fraction of the window's time-domain energy carried by stream s. Therefore, windows containing sharp or localized events typically concentrate energy in high-frequency detail streams, making $\mathbf{r}$ a useful indicator of when fine-scale is informative for time series composition. The $\mathbf{r}$ is leveraged as an auxiliary regularizer during training to calibrate the model's scale allocation, as specified in the pre-training objective.

2.5 Pre-training objectives

Masked Coefficient Reconstruction: In pre-training, a lightweight reconstruction head is applied per coefficient stream. For each stream $s \in \mathcal{S}$, the masked patch $\hat{\mathbf{p}}_{s,i}$ is reconstructed from the encoded token using a stream-specific head g_s, and unpatchify the patches to obtain reconstructed coefficients $\{\hat{\mathbf{c}}_s\}_{s\in\mathcal{S}}$. First, the reconstruction of the masked wavelet coefficients is supervised, providing scale-separated learning signals so that each scale is learned explicitly rather than only through an aggregated time-domain objective, $\mathcal{L}_{\text{coef}} = \sum_{s\in\mathcal{S}} \sum_{i=1}^{N} m_{s,i} \ \|\hat{\mathbf{p}}_{s,i} - \mathbf{p}_{s,i}\|_2^2$. This scale-separated supervision complements the bias of time-domain reconstruction toward low-frequency structure and improves sensitivity to fine-scale dynamics.

Time-domain Reconstruction: The reconstructed coefficient streams are composed via the inverse wavelet transform to obtain a time-domain reconstruction $\hat{\mathbf{x}}$, and minimize a reconstruction loss against the input time series: $\mathcal{L}_{\text{time}} = \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2$, where $\hat{\mathbf{x}} = \mathcal{W}_L^{-1}\big(\{\hat{\mathbf{c}}_s\}_{s\in\mathcal{S}}\big)$.

Scale Energy Ratio Distribution Calibration: Finally, the model's scale allocation is calibrated by aligning a predicted scale-importance distribution with the ground truth scale energy ratio distribution. Stream-level representations are obtained by taking the SST token outputs from each coefficient stream, $\mathbf{h}_{\mathcal{S}} = \big[\mathbf{h}_{s,0}^{\text{coef}}\big]_{s\in\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|\times D}$, and scale energy importance is predicted as $\hat{\mathbf{r}} = \mathrm{Softmax}(g_{er}(\mathbf{h}_{\mathcal{S}}))$, where g_{er} is a linear head. The KL-divergence $\mathcal{L}_{\text{ER}} = D_{\text{KL}}(\mathbf{r} \,\|\, \hat{\mathbf{r}})$ is then minimized. This calibration is particularly helpful for windows with localized events, where energy concentrates in high-frequency details, encouraging the model to preserve high-frequency details that time-domain reconstruction alone tends to smooth out. The overall pre-training objectives of WaveFM are $\mathcal{L} = \mathcal{L}_{\text{coef}} + \mathcal{L}_{\text{time}} + \mathcal{L}_{\text{ER}}$.

3 Experiments

We evaluate WaveFM against encoder-only TSFM baselines on two time series tasks, to test whether wavelet decomposed masked reconstruction yields transferable representations. For forecasting, we fine-tune all models with a lightweight forecasting head under different horizons. For imputation,

we evaluate WaveFM in a zero-shot setting, while MOMENT is fine-tuned on the pre-training head. Full settings on pre-training, model parameters, and downstream tasks are provided in Appendix C.

Long-horizon forecasting. Overall, WaveFM achieves consistently lower errors than encoder-only TSFM baselines across datasets, indicating that wavelet domain pre-training improves long-range predictive representations. The gains are particularly visible on datasets where fine-scale variations and regime changes are important, supporting our motivation that explicitly modeling multi-scale components benefits forecasting. Full experiment result is provided in Table 8.

Imputation. WaveFM shows strong zero-shot imputation performance, substantially improving over the TSFM baseline across diverse datasets. We attribute this to scale-separated coefficient reconstruction and cross-scale composition learning during pre-training, which encourages sensitivity to fine-grained dynamics that are often over-smoothed by the time-domain reconstruction objective. Notably, WaveFM remains competitive even in a zero-shot setting, suggesting that the learned representations transfer well to reconstruction without data-specific adaptation. Full experiment result is provided in Table 9.

4 Conclusion and Future Works

In this paper, we propose WaveFM, a wavelet domain multi-scale TSFM. WaveFM learns to understand how time series is decomposed into scale-separated wavelet streams and explicitly learns how these multi-scale components recompose into the original time series via the inverse wavelet transform. By combining masked coefficient reconstruction, scale coefficient attention, and an energy ratio distribution calibration, WaveFM provides a composition-aware self-supervisory signal that complements standard time-domain masked reconstruction. Across long-horizon forecasting and imputation benchmarks, WaveFM consistently improves over strong encoder-only TSFM baselines, supporting the importance of explicitly modeling multi-scale components for time series understanding and transferable representations.

For future work, we will further validate the time-series understanding learned by WaveFM through broader multi-task evaluations. Beyond long-horizon forecasting and imputation, we will extend our experiments to additional downstream tasks, including anomaly detection, short-term forecasting, and classification. We will also investigate how scale energy ratio distribution can be incorporated into downstream tasks as task-relevant signals, providing additional supervision or inductive bias tailored to each task.

REFERENCES

Alexander Alexandrov, Konstantinos Benidis, Michael Bohlke-Schneider, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Danielle C. Maddix, Syama Rangapuram, David Salinas, Jasper Schulz, Lorenzo Stella, Ali Caner Türkmen, and Yuyang Wang. Gluonts: Probabilistic and neural time series modeling in python. *Journal of Machine Learning Research*, 21(116):1–6, 2020. URL http://jmlr.org/papers/v21/19-820.html.

Abdul Fatir Ansari, Lorenzo Stella, Ali Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024.

Anthony Bagnall, Hoang Anh Dau, Jason Lines, Michael Flynn, James Large, Aaron Bostrom, Paul Southam, and Eamonn Keogh. The uea multivariate time series classification archive, 2018. *arXiv preprint arXiv:1811.00075*, 2018.

Yanlong Chen, Mattia Orlandi, Pierangelo Maria Rapa, Simone Benatti, Luca Benini, and Yawei Li. Physiowave: A multi-scale wavelet-transformer for physiological signal representation. *Advances in Neural Information Processing Systems*, 2025.

Mingyue Cheng, Qi Liu, Zhiding Liu, Hao Zhang, Rujiao Zhang, and Enhong Chen. Timemae: Self-supervised representations of time series with decoupled masked autoencoders. *arXiv preprint arXiv:2303.00320*, 2023.

Charles R Cornish, Christopher S Bretherton, and Donald B Percival. Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. *Boundary-Layer Meteorology*, 119(2):339–374, 2006.

Hoang Anh Dau, Eamonn Keogh, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, Yanping, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen, Gustavo Batista, and Hexagon-ML. The ucr time series classification archive, October 2018. https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.

Patrick Emami, Abhijeet Sahu, and Peter Graf. Buildingsbench: A large-scale dataset of 900k buildings and benchmark for short-term load forecasting. *Advances in Neural Information Processing Systems*, 36:19823–19857, 2023.

Rakshitha Godahewa, Christoph Bergmeir, Geoffrey I. Webb, Rob J. Hyndman, and Pablo Montero-Manso. Monash time series forecasting archive. In *Neural Information Processing Systems Track on Datasets and Benchmarks*, 2021.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. Moment: a family of open time-series foundation models. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 16115–16152, 2024.

Haokun Gui, Xiucheng Li, and Xinyang Chen. Vector quantization pretraining for eeg time series with random projection and phase alignment. In *International Conference on Machine Learning*, pp. 16731–16750. PMLR, 2024.

Tiantian Guo, Tongpo Zhang, Enggee Lim, Miguel López-Benítez, Fei Ma, and Limin Yu. A review of wavelet analysis and its applications: Challenges and opportunities. *IEEE Access*, 10:58869–58903, 2022. doi: 10.1109/ACCESS.2022.3179517.

Yudong Han, Haocong Wang, Yupeng Hu, Yongshun Gong, Xuemeng Song, and Weili Guan. Content-aware balanced spectrum encoding in masked modeling for time series classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17059–17067, 2025.

Jiawei Jiang, Chengkai Han, Wenjun Jiang, Wayne Xin Zhao, and Jingyuan Wang. Libcity: A unified library towards efficient and comprehensive urban spatial-temporal prediction. *arXiv preprint arXiv:2304.14343*, 2023.

Gregory Lee, Ralf Gommers, Filip Waselewski, Kai Wohlfahrt, and Aaron O'Leary. Pywavelets: A python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, 2019.

Jae-Neung Lee, Myung-Won Lee, Yeong-Hyeon Byeon, Won-Sik Lee, and Keun-Chang Kwak. Classification of horse gaits using fcm-based neuro-fuzzy classifier from the transformed data information of inertial sensor. *Sensors*, 16(5):664, 2016.

Beibu Li, Qichao Shentu, Yang Shu, Hui Zhang, Ming Li, Ning Jin, Bin Yang, and Chenjuan Guo. Crossad: Time series anomaly detection with cross-scale associations and cross-window modeling. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Zhe Li, Zhongwen Rao, Lujia Pan, Pengyun Wang, and Zenglin Xu. Ti-mae: Self-supervised masked time series autoencoders. *arXiv preprint arXiv:2301.08871*, 2023.

Yong Liu, Haoran Zhang, Chenyu Li, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Timer: generative pre-trained transformers are large time series models. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 32369–32399, 2024.

S.G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989. doi: 10.1109/34.192463.

Luca Masserano, Abdul Fatir Ansari, Boran Han, Xiyuan Zhang, Christos Faloutsos, Michael W Mahoney, Andrew Gordon Wilson, Youngsuk Park, Syama Sundar Rangapuram, Danielle C Maddix, et al. Enhancing foundation models for time series forecasting via wavelet-based tokenization. In *Forty-second International Conference on Machine Learning*, 2024.

Md Mahmuddun Nabi Murad, Mehmet Aktukmak, and Yasin Yilmaz. Wpmixer: Efficient multi-resolution mixing for long-term time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 19581–19588, 2025.

John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S Tsay, Themis Palpanas, and Michael J Franklin. Tsb-uad: an end-to-end benchmark suite for univariate time-series anomaly detection. *Proceedings of the VLDB Endowment*, 15(8):1697–1711, 2022.

Harshavardhan Prabhakar Kamarthi and B Aditya Prakash. Large pre-trained time series models for cross-domain time series analysis tasks. *Advances in Neural Information Processing Systems*, 37: 56190–56214, 2024.

Xiaoming Shi, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen, and Ming Jin. Time-moe: Billion-scale time series foundation models with mixture of experts. In *The Thirteenth International Conference on Learning Representations*, 2024.

Ariel Slepyan, Michael Zakariaie, Trac Tran, and Nitish Thakor. Wavelet transforms significantly sparsify and compress tactile interactions. *Sensors*, 24(13):4243, 2024.

Christopher Torrence and Gilbert P Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1):61–78, 1998.

Wenxuan Wang, Kai Wu, Yujian Betterest Li, Dan Wang, and Xiaoyu Zhang. Synthetic series-symbol data generation for time series foundation models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025a.

Yubo Wang, Hui He, Chaoxi Niu, and Zhendong Niu. Wavetuner: Comprehensive wavelet subband tuning for time series forecasting. *arXiv preprint arXiv:2511.18846*, 2025b.

Gerald Woo, Chenghao Liu, Akshat Kumar, and Doyen Sahoo. Pushing the limits of pre-training for time series forecasting in the cloudops domain. *arXiv preprint arXiv:2310.05063*, 2023.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 53140–53164, 2024.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34:22419–22430, 2021.

Shifeng Xie, Vasilii Feofanov, Marius Alonso, Ambroise Odonnat, Jianfeng Zhang, Themis Palpanas, and Ievgen Redko. Cauker: classification time series foundation models can be pretrained on synthetic data only. *arXiv preprint arXiv:2508.02879*, 2025.

Xiaoxiao Yang, Xiaojiao Chen, Kezhong Sun, Chuanyu Xiong, Dongran Song, Yingchun Lu, Liansheng Huang, Shiying He, and Xiuqing Zhang. A wavelet transform-based real-time filtering algorithm for fusion magnet power signals and its implementation. *Energies*, 16(10):4091, 2023.

Haoran Zhang, Yong Liu, Yunzhong Qiu, Haixuan Liu, Zhongyi Pei, Jianmin Wang, and Mingsheng Long. Timesbert: A bert-style foundation model for time series understanding. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 10975–10983, 2025.

A WAVELET TRANSFORM

Wavelets provide a multi-resolution view of a signal by decomposing it into one coarse component and a sequence of residual components at progressively finer temporal scales (Mallat, 1989). In our setting, an L-level discrete wavelet transform (DWT) maps an input window $\mathbf{x} \in \mathbb{R}^T$ to an approximation coefficient stream $\mathbf{a}^{(L)}$ and detail coefficient streams $\{\mathbf{d}^{(\ell)}\}_{\ell=1}^{L}$ (Figure 3), where $\mathbf{a}^{(L)}$ summarizes the low-frequency trend and each $\mathbf{d}^{(\ell)}$ captures scale-specific fluctuations relative to the coarser representation. These coefficient streams are the multi-scale inputs used by WaveFM, as described in the main paper.

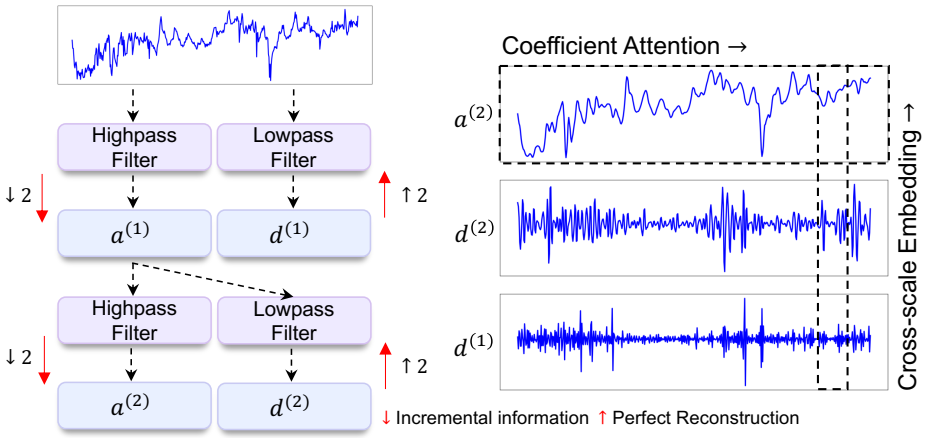

Figure 3: Wavelet transform. The multi-level discrete wavelet transform decomposes the input time series into one low-frequency approximation stream and multiple high-frequency detail streams, providing scale-separated representations.

Forward transform $\mathcal{W}_L$. Let $h[\cdot]$ and $g[\cdot]$ denote the low-pass and high-pass analysis filters associated with a chosen wavelet family, and define the level-0 approximation coefficients as $\mathbf{a}^{(0)} = \mathbf{x}$. For each level $\ell = 1, \ldots, L$, the DWT applies filtering followed by dyadic downsampling to produce the next approximation stream and the corresponding detail stream:

$$a^{(\ell)}[k] = \sum_n h[n-2k]\, a^{(\ell-1)}[n], \qquad d^{(\ell)}[k] = \sum_n g[n-2k]\, a^{(\ell-1)}[n].$$

The resulting coefficient set is

$$\mathcal{W}_L(\mathbf{x}) = \left(\mathbf{a}^{(L)}, \{\mathbf{d}^{(\ell)}\}_{\ell=1}^{L}\right), \qquad \mathcal{S} = \{\mathbf{a}^{(L)}, \mathbf{d}^{(L)}, \ldots, \mathbf{d}^{(1)}\}.$$

Choice of wavelet basis. In our implementation, we use symlet-4 (`sym4`) as the default wavelet basis. Symlets offer a favorable trade-off between time localization and frequency selectivity via moderately long, near-symmetric filters, which helps reduce phase distortion and boundary artifacts in discrete-time signals(Slepyan et al., 2024). Compared to shorter bases such as haar, sym4 provides smoother approximation streams and more stable detail coefficients, while remaining computationally lightweight compared to higher-order bases(Cornish et al., 2006). Empirically, this choice yields robust multi-scale coefficient representations across diverse time series domains without introducing excessive filter length or sensitivity to edge effects (Guo et al., 2022).

Incremental information across levels. The multi-level construction of $\mathcal{W}_L$ can be viewed as an incremental information process; at each level ℓ, the approximation $\mathbf{a}^{(\ell)}$ provides a progressively coarser summary of $\mathbf{x}$, while the detail coefficients $\mathbf{d}^{(\ell)}$ capture the information discarded by the low-pass downsampled approximation when moving from level $\ell - 1$ to ℓ(Lee et al., 2016). Intuitively, $\mathbf{d}^{(\ell)}$ encodes the residual information needed to refine $\mathbf{a}^{(\ell)}$ back toward $\mathbf{a}^{(\ell-1)}$. Collecting details across levels, therefore, yields a hierarchy of residual streams, from coarse-to-fine, that together characterize temporal dynamics at multi-scale.

Coefficient lengths and scale interpretation. Each level downsamples by a factor of 2, so the temporal resolution decreases with ℓ. With common boundary extension rules, coefficient lengths scale on the order of $T/2^\ell$:

$$|\mathbf{a}^{(\ell)}| \approx \left\lceil \frac{T}{2^\ell} \right\rceil, \qquad |\mathbf{d}^{(\ell)}| \approx \left\lceil \frac{T}{2^\ell} \right\rceil,$$

where exact lengths depend on the extension mode and the wavelet filter length (Lee et al., 2019). Thus, $\mathbf{a}^{(L)}$ represents the coarsest trend over the entire window, while $\mathbf{d}^{(1)}$ captures the finest-scale residual variations. This motivates using scale-dependent patch lengths P_s in WaveFM so that each stream produces the same number of patches N despite different native resolutions.

Inverse transform $\mathcal{W}_L^{-1}$ and perfect reconstruction. Let $\tilde{h}[\cdot]$ and $\tilde{g}[\cdot]$ denote the low-pass and high-pass synthesis filters. For $\ell = L, \ldots, 1$, the inverse recursion reconstructs $\mathbf{a}^{(\ell-1)}$ from $\mathbf{a}^{(\ell)}$ and $\mathbf{d}^{(\ell)}$ by dyadic upsampling and filtering:

$$a^{(\ell-1)}[n] = \sum_k \tilde{h}[n-2k]\; a^{(\ell)}[k] + \sum_k \tilde{g}[n-2k]\; d^{(\ell)}[k].$$

Applying this recursion from $\ell = L$ down to $\ell = 1$ yields the inverse mapping

$$\mathbf{x} = \mathcal{W}_L^{-1}\Big(\mathbf{a}^{(L)}, \{\mathbf{d}^{(\ell)}\}_{\ell=1}^{L}\Big).$$

If $(h, g, \tilde{h}, \tilde{g})$ form a perfect reconstruction filter bank and the same extension rule is used in analysis and synthesis, then $\mathcal{W}_L^{-1}\big(\mathcal{W}_L(\mathbf{x})\big) = \mathbf{x}$ holds up to numerical precision.

Boundary handling. Because filtering requires samples beyond the finite window, practical DWT implementations adopt a boundary extension rule (periodic or symmetric extension) (Yang et al., 2023). In WaveFM, $\mathcal{W}_L$ and $\mathcal{W}_L^{-1}$ use the same boundary handling so that reconstruction remains stable and coefficient streams align consistently across windows.

B Related Works

B.1 Encoder-only TSFMs

Encoder-only TSFMs typically follow a BERT-style pre-training paradigm (Gui et al., 2024), where a Transformer encoder learns transferable representations by reconstructing masked portions of a time series or its patch tokens from the surrounding bidirectional context (Cheng et al., 2023). This enables strong adaptation to diverse downstream tasks. TimesBERT follows this encoder-only design for multivariate time series and pre-trains with masked patch modeling and additional objectives to improve transfer across datasets and tasks (Zhang et al., 2025). Ti-MAE uses masked autoencoding over tokenized time series and shows that reconstruction-based pre-training can learn useful representations for forecasting and classification while reducing the gap between pre-training and downstream objectives (Li et al., 2023). LPTM further explores pre-training with adaptive segmentation so that the model can form consistent tokens across different datasets, improving transfer under fine-tuning and in zero-shot settings (Prabhakar Kamarthi & Prakash, 2024). Together, these results support encoder-only backbones as efficient foundation models for time series representation learning when broad transfer is preferred over autoregressive generation. However, time-domain masked reconstruction fails to explicitly disentangle scale-specific information and supervise cross-scale interactions required to compose the original signal. We address this problem with wavelet domain masked coefficient reconstruction. Scale-coefficient attention (SCA) aggregates information across scales and time, and complements masked coefficient reconstruction by modeling how multi-scale components jointly compose the original time series.

B.2 Wavelet Transform on Time Series Analysis

Wavelet transform is a classical tool for time-series analysis providing multi-resolution time-frequency localization, enabling signals with temporal structure to be decomposed into approximation and detail coefficients for scale-wise inspection (Torrence & Compo, 1998). Recent time series analysis models adopt this property in different ways. For forecasting, WaveTuner performs wavelet

decomposition and learns an adaptive router that assigns importance weights to sub-bands, explicitly emphasizing informative frequency components to improve forecasting performance (Wang et al., 2025b). Similarly, WPMixer decomposes an input series and processes each resolution in a dedicated branch, leveraging multi-resolution mixing to enhance long-horizon forecasting (Murad et al., 2025). In a foundation model approach, WaveToken uses a decimated DWT as a tokenizer by quantizing wavelet coefficients and pretraining an autoregressive model directly in the space of time-localized frequencies, targeting general-purpose forecasting (Masserano et al., 2024). For representation learning on biosignals, PhysioWave introduces a learnable wavelet decomposition pipeline to adaptively capture heterogeneous, non-stationary physiological dynamics during self-supervised pretraining (Chen et al., 2025). While these works demonstrate that wavelet transforms are useful intermediate representations, most of them optimize task-driven objectives such as forecasting; however, they do not explicitly model how within-scale structure and cross-scale interactions compose the time series. Our approach addresses this gap by pre-training with masked coefficient reconstruction coupled with explicit inverse wavelet transform, thereby encouraging the model to learn how multi-scale components jointly explain the time series.

C Experiment settings

C.1 Pre-training datasets

WaveFM is pre-trained on a diverse set of public time-series datasets, including subsets from Monash (Godahewa et al., 2021), UEA/UCR (Bagnall et al., 2018; Dau et al., 2018), and TSB-UAD (Paparrizos et al., 2022). Also, we include datasets provided by previously curated large-scale collections, UTSD (Liu et al., 2024), LOTSA (Woo et al., 2024), and Time Series Pile (Goswami et al., 2024), along with selected subsets from BuildingsBench (Emami et al., 2023), LibCity (Jiang et al., 2023), CloudOps (Woo et al., 2023), GluonTS (Alexandrov et al., 2020), and some other time series datasets (Ansari et al., 2024). In total, WaveFM is pre-trained on **29.18 GB** of data, comprising **2.61B** observations (timestamps $\times$ channels), as summarized in Tables 3–5. Following prior work, we convert multivariate series into channel-wise univariate sequences under a channel-independent setting (Shi et al., 2024; Xie et al., 2025).

Table 3: Pre-training datasets from Paparrizos et al. (2022).WaveFM is pre-trained on the training split of each dataset.

Dataset	# Channels	# Obs.
ECG	1	7,186,928
KDD21	1	4,545,977
IOPS	1	1,802,408
YAHOO	1	343,451
NAB	1	219,249
NASA-MSM	1	118,215
NASA-MSL	1	49,949
Occupancy	1	14,655

C.2 Pre-training settings

Pre-training uses the datasets in Appendix C.1 with fixed-length windows of 512 and a stride of 8. Each window is patchified into non-overlapping patches of length 8 (stride 8). We apply a 3-level discrete wavelet transform with the Symlet-4 basis (sym4), producing one approximation stream and three detail streams. For every wavelet coefficient stream, 30% of patches are randomly masked, and WaveFM is trained to reconstruct the missing coefficients; the reconstructed coefficients are then mapped back to the time domain via the inverse wavelet transform. All WaveFM variants are trained for 2 epochs, with architectural configurations, such as encoder layers, hidden size, and attention setup, summarized in Table 6.

Table 4: Pre-training datasets. We report the total number of time stamps (including channels).

Dataset	# Obs.	Dataset	# Obs.
Residential_Load_Power	437,983,677	Health_PigArtPressure	624,000
Residential_Pv_Power	376,016,850	Health_PigCVP	624,000
Q-TRAFFIC	264,386,688	SZ_TAXI	464,256
Alibaba_Cluster_Trace_2018	190,385,060	M5	457,350
Wind_Farms_Minutely	172,178,060	Bdg-2_Hog	421,056
London_Smart_Meters	166,528,896	Godaddy	257,070
IoT_Baian	165,361,266	Nature_Worms	232,200
Kaggle_Web_Traffic	116,485,589	Hierarchical_Sales	212,164
Health_TDBrain_Csv	73,299,996	Cdc_Fluview_Who_Nrevss	167,040
Health_MotorImagery	72,576,000	Gfc17_Load	140,352
Health_BIDMC32HR	63,592,000	Car_Parts	136,374
Subseasonal	56,788,560	Tourism_Monthly	109,280
Taxi_30min	54,999,056	Smart	95,709
Weather	43,032,000	Traffic_Weekly	89,648
Wiki-rolling_Nips	40,619,100	Nn5_Daily	87,801
LOOP_SEATTLE	33,953,760	Borealis	83,269
Nature_EigenWorms	27,947,136	Fred_Md	77,896
PEMS07	24,921,792	Bitcoin	75,364
Temperature_Rain	23,252,200	Sunspot	73,924
Dominick	19,092,987	Hospital	64,428
Environment_BenzeneConcentration	16,343,040	Covid_Deaths	56,392
PEMS04	15,649,632	M1_Monthly	55,998
Health_IEEEPPG	15,480,000	Electricity_Weekly	50,076
Traffic_Hourly	15,122,928	Uber_Tlc_Daily	47,087
Environment_AustraliaRainfall	11,539,224	Tourism_Quarterly	42,544
Subseasonal_Precip	9,760,426	Vehicle_Trips	42,382
Nature_StarLightCurves	9,457,664	Health_AtrialFibrillation	38,400
PEMS03	9,382,464	Spain	35,064
PEMS08	9,106,560	Sceaux	34,223
Electricity_Hourly	8,443,584	Covid19_Energy	31,912
Solar_4_Seconds	7,397,222	Saugeenday	23,741
Energy_Wind_4_Seconds	7,397,147	Elf	21,792
Solar_10_Minutes	7,200,720	Bdg-2_Cockatoo	17,544
LOS_LOOP	7,094,304	Gfc14_Load	17,520
Bdg-2_Rat	4,728,288	Elecdemand	17,520
Environment_BeijingPM25Quality	3,664,656	Pdb	17,520
Pedestrian_Counts	3,132,346	M3_Other	13,325
Health_SelfRegulationSCP2	3,064,320	Tourism_Yearly	12,757
Health_SelfRegulationSCP1	3,015,936	Nn5_Weekly	12,543
Kdd_Cup_2018	2,942,364	M1_Quarterly	9,944
Bdg-2_Fox	2,324,568	Us_Births	7,305
Nature_Phoneme	2,160,640	Solar_Weekly	7,124
Bdg-2_Bear	1,482,312	Cif_2016	7,108
Ideal	1,255,253	M1_Yearly	4,515
Rideshare	1,246,464	HeartRate	1,801
Energy_Australian_Electricity_Demand	1,155,264	GasRateCO2	593
Uber_Tlc_Hourly	1,129,444	AusBeer	217
Bdg-2_Panther	919,800	Wine	182
Oikolab_Weather	800,456	MonthlyMilk	174
Gfc12_Load	788,280	AirPassengers	150
Bdg-2_Bull	719,304	Wooly	125

C.3 Fine-tuning settings

We keep the wavelet basis and decomposition level fixed to the pre-training configuration. We also report parameter sizes for each model in Table 7. Compared to MOMENT-base and LPTM, WaveFM-small and WaveFM-base use substantially fewer parameters, while achieving competitive performance.

Long-horizon Forecasting. We fine-tune the pre-trained WaveFM encoder for downstream tasks. For long-horizon forecasting, we attach a lightweight forecasting head for each wavelet stream, and train only the heads while keeping the encoder frozen. Each stream-specific head predicts the future wavelet coefficients for its corresponding stream, and the final forecast in the time domain is obtained by inverse wavelet transformation. To ensure a fair comparison, we freeze the pre-

Table 5: Classification datasets from Dau et al. (2018); Bagnall et al. (2018). We train on the training split of each dataset (including channels).

Dataset	# Ch.	Time series Length	# Obs.
InsectWingbeat	200	22	94,283,200
MotorImagery	64	3,000	45,696,000
FaceDetection	144	62	45,068,544
PEMS-SF	963	144	31,617,216
DuckDuckGeese	1,345	270	15,252,300
EigenWorms	6	17,984	11,761,536
SpokenArabicDigits	13	93	6,838,104
PhonemeSpectra	11	217	6,781,467
Heartbeat	61	405	4,298,670
HandOutlines	1	2,709	2,321,613
FordB	1	500	1,558,000
FordA	1	500	1,543,000
SelfRegulationSCP2	7	1,152	1,378,944
SelfRegulationSCP1	6	896	1,231,104
EthanolConcentration	3	1,751	1,171,419
NonInvasiveFetalECGThorax1	1	750	1,156,500
NonInvasiveFetalECGThorax2	1	750	1,156,500
StarLightCurves	1	1,024	877,568
EthanolLevel	1	1,751	756,432
ElectricDevices	1	96	734,400
UWaveGestureLibraryAll	1	945	725,760
Cricket	6	1,197	660,744
PLAID	1	1,344	618,240
SemgHandSubjectCh2	1	1,500	577,500
SemgHandMovementCh2	1	1,500	577,500
HandMovementDirection	10	400	548,000
LSST	6	36	455,112
MixedShapesRegularTrain	1	1,024	438,272
EOGHorizontalSignal	1	1,250	387,500
EOGVerticalSignal	1	1,250	387,500
SemgHandGenderCh2	1	1,500	385,500
FingerMovements	28	50	378,000
ArticularyWordRecognition	9	144	304,560
Crop	1	46	283,866
ShapesAll	1	512	263,168
UWaveGestureLibraryZ	1	315	241,920
UWaveGestureLibraryY	1	315	241,920
UWaveGestureLibraryX	1	315	241,920
LargeKitchenAppliances	1	720	231,120
ScreenType	1	720	231,120
SmallKitchenAppliances_Industrial	1	720	231,120
SmallKitchenAppliances	1	720	231,120
RefrigerationDevices	1	720	231,120
NATOPS	24	51	188,496
Phoneme	1	1,024	187,392
PigAirwayPressure	1	2,000	178,000
PigArtPressure	1	2,000	178,000
PigCVP	1	2,000	178,000
InlineSkate	1	1,882	159,970
Computers	1	720	154,080
Haptics	1	1,092	144,144
Earthquakes	1	512	141,312
Worms	1	900	139,500
WormsTwoClass	1	900	139,500
Wafer	1	152	130,264
ACSF1	1	1,460	124,100
Strawberry	1	235	123,375
PhalangesOutlinesCorrect	1	80	123,360
TwoPatterns	1	128	109,696
Yoga	1	426	109,482
FiftyWords	1	270	103,950
PenDigits	2	8	102,768
CricketY	1	300	100,200
CricketZ	1	300	100,200
CricketX	1	300	100,200
StandWalkJump	4	2,500	100,000
AllGestureWiimoteX	1	385	98,945
UWaveGestureLibrary	3	315	96,390
AllGestureWiimoteY	1	369	94,833
MixedShapesSmallTrain	1	1,024	87,040
AllGestureWiimoteZ	1	326	83,782
OSULeaf	1	427	73,017
Epilepsy	3	206	72,306
JapaneseVowels	12	26	72,072
Fish	1	463	69,450
HouseTwenty	1	2,000	68,000
ChlorineConcentration	1	166	66,400
GestureMidAirD1	1	360	64,080
GestureMidAirD2	1	360	64,080
GestureMidAirD3	1	360	64,080

Dataset	# Ch.	Time series Length	# Obs.
FaceAll	1	131	62,880
WordSynonyms	1	270	61,560
ECG5000	1	140	59,920
Adiac	1	176	58,784
Handwriting	3	152	58,368
GesturePebbleZ2	1	455	56,875
CinCECGTorso	1	1,639	55,726
SwedishLeaf	1	128	54,784
GesturePebbleZ1	1	455	51,415
Rock	1	2,844	48,348
Mallat	1	1,024	48,128
InsectWingbeatSound	1	256	48,128
DistalPhalanxOutlineCorrect	1	80	41,120
ProximalPhalanxOutlineCorrect	1	80	41,120
MiddlePhalanxOutlineCorrect	1	80	41,120
Ham	1	431	40,083
FreezerRegularTrain	1	301	38,528
Lightning2	1	637	32,487
MedicalImages	1	99	32,274
InsectEPGRegularTrain	1	601	31,853
Car	1	577	29,427
Herring	1	512	27,648
DistalPhalanxTW	1	80	27,360
MiddlePhalanxOutlineAgeGroup	1	80	27,360
ProximalPhalanxTW	1	80	27,360
ProximalPhalanxOutlineAgeGroup	1	80	27,360
MiddlePhalanxTW	1	80	27,360
DistalPhalanxOutlineAgeGroup	1	80	27,360
MelbournePedestrian	1	24	24,552
Trace	1	275	23,375
RacketSports	6	30	23,220
Meat	1	448	22,848
FacesUCR	1	131	22,401
PowerCons	1	144	22,176
BasicMotions	6	100	20,400
Lightning7	1	319	19,140
DodgerLoopDay	1	288	19,008
GunPointOldVersusYoung	1	150	17,400
GunPointAgeSpan	1	150	17,250
GunPointMaleVersusFemale	1	150	17,250
ShakeGestureWiimoteZ	1	385	16,170
SyntheticControl	1	60	15,420
AtrialFibrillation	2	640	15,360
PickupGestureWiimoteZ	1	361	15,162
OliveOil	1	570	14,250
Libras	2	45	13,860
Plane	1	144	12,960
Beef	1	470	11,750
Wine	1	234	11,232
ToeSegmentation2	1	343	10,290
DodgerLoopWeekend	1	288	9,792
ToeSegmentation1	1	277	9,418
BirdChicken	1	512	8,704
BeetleFly	1	512	8,704
ShapeletSim	1	500	8,500
InsectEPGSmallTrain	1	601	8,414
Symbols	1	398	8,358
ECG200	1	96	8,160
ArrowHead	1	251	7,530
FreezerSmallTrain	1	301	7,224
FaceFour	1	350	7,000
Coffee	1	286	6,864
ERing	4	65	6,500
GunPoint	1	150	6,300
DodgerLoopGame	1	288	4,896
UMD	1	150	4,500
DiatomSizeReduction	1	345	4,485
BME	1	128	3,200
CBF	1	128	3,200
Fungi	1	201	3,015
ECGFiveDays	1	136	2,584
SmoothSubspace	1	15	1,920
TwoLeadECG	1	82	1,558
SonyAIBORobotSurface2	1	65	1,495
MoteStrain	1	84	1,428
ItalyPowerDemand	1	24	1,368
SonyAIBORobotSurface1	1	70	1,190
Chinatown	1	24	408

Table 6: WaveFM architectural configurations.

	WaveFM-small	**WaveFM-base**
Encoder layers	3	6
Hidden size (d_{model})	256	512
Attention heads	4	8
FFN size (d_{ff})	1024	2048
Dropout	0.1	0.1

trained encoders for all baselines and train only their forecasting heads, while following the official hyperparameter setting for each baselines.

Imputation. For imputation, we evaluate WaveFM in a zero-shot setting using the pre-training reconstruction head, without any dataset-specific fine-tuning. We randomly mask a subset of time steps and the corresponding wavelet streams. Then, we reconstruct each wavelet coefficient from the observed context. For baseline, we freeze the pre-trained encoder and train only its task-specific head, following the official hyperparameter setting.

Table 7: Model size comparison in terms of the total number of parameters (in millions).

Model	Params (M)
MOMENT-base	109.64
LPTM	109.71
WaveFM-small	12.17
WaveFM-base	47.93

D VISUALIZATION

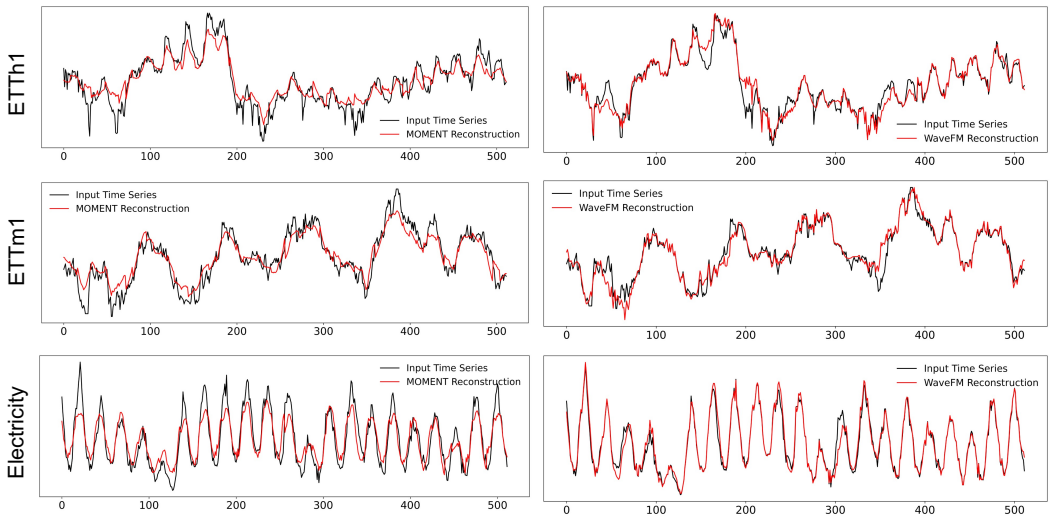

Figure 4: Reconstruction visualization. We visualize reconstruction outputs on three datasets (ETTh1, ETTm1, Electricity). The input time series is shown in black, and the model reconstruction is shown in red. The left column reports MOMENT reconstructions, and the right column reports WaveFM reconstructions.

Figure 4 provides qualitative evidence of how different pre-training objectives affect reconstruction behavior. Across ETTh1 and ETTm1, both models broadly follow the global trajectory, but clear differences emerge in local fluctuations and abrupt changes. On Electricity, which exhibits a strong

periodic structure with sharp peaks, the reconstructions highlight how well each model preserves oscillatory patterns and peak timing under the same input window. Overall, WaveFM learns how information across temporal scales complements to compose the original time series with masked coefficient reconstruction. It leverages coarse-scale context to disambiguate local variations while using fine-scale cues to recover short transitions and transient deviations that are often smoothed out when only using time-domain masked reconstruction.

E FULL RESULTS

Table 8: Long-horizon forecasting results.

Dataset	Horizon	WaveFM-small		WaveFM-base		MOMENT-base		LPTM	
		MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
ETTh1	96	**0.4074**	**0.3891**	0.4145	0.3948	0.4206	0.3973	0.4428	0.4367
	192	**0.4238**	**0.4123**	0.4275	0.4136	0.4364	0.4221	0.4542	0.4564
	336	0.4497	0.4332	**0.4437**	**0.4265**	0.4537	0.4456	0.4700	0.4695
	720	0.4888	0.4755	**0.4639**	**0.4406**	0.5010	0.5118	0.5090	0.5335
ETTh2	96	**0.3493**	0.2942	0.3518	**0.2932**	0.3624	0.3093	0.3654	0.3240
	192	**0.3826**	0.3452	0.3835	**0.3423**	0.3948	0.3585	0.3968	0.3682
	336	0.4009	0.3612	**0.4007**	**0.3581**	0.4060	0.3642	0.4114	0.3754
	720	0.4350	0.3963	**0.4274**	**0.3848**	0.4393	0.4018	0.4381	0.4023
ETTm1	96	0.3620	0.3257	**0.3604**	0.3215	0.3674	0.3276	0.3622	**0.3190**
	192	0.3785	0.3545	0.3769	0.3505	0.3849	0.3605	**0.3764**	**0.3464**
	336	0.3932	0.3816	0.3920	0.3778	0.4009	0.3904	**0.3912**	**0.3732**
	720	0.4206	0.4316	0.4209	0.4305	0.4278	0.4405	**0.4176**	**0.4227**
ETTm2	96	0.2633	0.1752	**0.2594**	**0.1714**	0.2716	0.1824	0.2664	0.1795
	192	0.2966	0.2254	**0.2936**	**0.2223**	0.3046	0.2300	0.2989	0.2282
	336	0.3287	0.2744	**0.3268**	**0.2728**	0.3345	0.2782	0.3342	0.2797
	720	0.3823	0.3596	**0.3815**	**0.3594**	0.3878	0.3603	0.3871	0.3615
Electricity	96	0.2340	**0.1370**	**0.2339**	**0.1370**	0.2782	0.1724	0.2484	0.1466
	192	0.2464	0.1514	**0.2461**	**0.1512**	0.2877	0.1842	0.2605	0.1606
	336	0.2623	**0.1669**	**0.2619**	**0.1669**	0.3004	0.1979	0.2752	0.1755
	720	0.2949	**0.2060**	**0.2942**	0.2061	0.3285	0.2345	0.3054	0.2130
Weather	96	**0.2215**	**0.1679**	0.2233	0.1691	0.2351	0.1779	0.2242	0.1685
	192	0.2582	0.2119	0.2592	0.2126	0.2678	0.2192	**0.2569**	**0.2082**
	336	0.2926	0.2577	0.2932	0.2584	0.2996	0.2635	**0.2911**	**0.2535**
	720	0.3390	0.3229	0.3400	0.3243	0.3459	0.3283	**0.3380**	**0.3198**

Long-horizon forecasting. Table 8 shows the full results of long-horizon forecasting. Both WaveFM variants (small, base) achieve competitive performance despite a substantial difference in the number of parameters, indicating that wavelet-domain pre-training yields parameter-efficient representations for long-horizon forecasting. Moreover, training only with a lightweight forecasting head suggests that the frozen encoder representations already encode multi-scale structures of future long-horizon dynamics in the wavelet coefficient streams. This is likely because scale coefficient attention (SCA) captures cross-scale dependencies, allowing information to flow across temporal scales.

Imputation. Table 9 shows the full results of imputation. Across all datasets and masking ratios, WaveFM achieves substantially lower errors than MOMENT, demonstrating strong robustness under sparse observations. This advantage continues as the missing rate increases, suggesting that masked reconstruction of wavelet coefficient streams provides informative multi-scale context for recovering missing values by leveraging both cross-scale and within-scale dependencies. Since WaveFM is evaluated in a zero-shot setting without any dataset-specific fine-tuning, these results indicate that the pre-trained encoder learns transferable multi-scale representations of time series.

Table 9: Imputation results.

Dataset	Ratio	WaveFM-small		WaveFM-base		MOMENT-base	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	12.5%	**0.068**	0.094	0.070	0.096	0.393	0.416
	25.0%	0.129	0.161	0.148	0.163	0.397	0.419
	37.5%	**0.193**	0.221	0.226	**0.220**	0.406	0.425
	50.0%	**0.302**	0.294	0.312	0.282	0.424	0.433
	Avg	**0.173**	0.192	0.189	**0.190**	0.405	0.423
ETTm2	12.5%	0.023	0.055	0.025	0.058	0.138	0.251
	25.0%	0.042	0.091	0.050	0.096	0.140	0.253
	37.5%	0.066	0.129	0.070	0.130	0.144	0.256
	50.0%	0.101	0.173	0.092	0.165	0.148	0.260
	Avg	0.058	0.112	0.059	0.112	0.143	0.255
ETTh1	12.5%	0.087	0.110	0.099	0.125	0.693	0.579
	25.0%	0.170	0.185	0.156	0.186	0.694	0.580
	37.5%	0.293	0.262	0.256	0.250	0.699	0.581
	50.0%	0.388	0.337	0.313	0.306	0.706	0.584
	Avg	0.234	0.223	0.206	0.217	0.698	0.581
ETTh2	12.5%	0.028	**0.064**	**0.028**	0.067	0.284	0.368
	25.0%	0.055	0.111	**0.052**	**0.107**	0.285	0.369
	37.5%	0.086	0.152	**0.079**	**0.145**	0.287	0.371
	50.0%	0.127	0.203	**0.103**	**0.185**	0.290	0.373
	Avg	0.074	0.132	**0.065**	**0.126**	0.287	0.370
Electricity	12.5%	0.068	0.123	0.073	0.129	0.167	0.298
	25.0%	0.137	0.209	0.144	0.207	0.174	0.303
	37.5%	0.201	0.279	0.185	**0.258**	**0.184**	0.309
	50.0%	0.311	0.373	0.222	**0.304**	**0.200**	0.319
	Avg	0.179	0.246	**0.156**	**0.224**	0.181	0.307
Weather	12.5%	**0.018**	0.044	0.021	0.045	0.090	0.154
	25.0%	**0.039**	**0.076**	0.046	0.080	0.093	0.155
	37.5%	**0.061**	0.106	0.072	0.110	0.098	0.158
	50.0%	0.091	0.139	**0.086**	0.135	0.103	0.161
	Avg	**0.052**	0.091	0.056	0.092	0.096	0.157

F ACKNOWLEDGMENTS

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (RS-2024-00443780, Development of Foundation Models for Bioelectrical Signal Data and Validation of Their Clinical Applications: A Noise-and-Variability Robust, Generalizable Self-Supervised Learning Approach; RS-2024-00439932, SW Starlab), Korean Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (RS-2024-00435815, Human Resource Development Program for Industrial Innovation(Global)), the Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (RS-2024-00463188), and the Korea Planning & Evaluation Institute of Industrial Technology (KEIT) funded by the Ministry of Trade, Industry and Resources (RS-2025-25458052, Development of Core Technologies for Manufacturing Foundation Models).